DOCUMENT RESUME

ED 256 039 EA 017 664

AUTHOR Conyers, John G.; And Others
TITLE Developing District Made Criterion Referenced Tests: A Standard of Excellence for Effective Schools.
INSTITUTION Denver Public Schools, Colo.
PUB DATE 84
NOTE 20p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Criterion Referenced Tests; *Educational Diagnosis; Elementary Secondary Education; Program Descriptions; *Standardized Tests; Student Evaluation; Test Construction; *Testing Programs; *Test Validity
IDENTIFIERS *Academic Learning Progress Assessment Program; Denver Public Schools CO

ABSTRACT

Research on the lack of correlation between standardized tests and school district goals has suggested that district goals might be better met by supplementing standardized tests with district-made criterion-referenced testing systems. This article reports on the development of such a system, the Academic Learning Progress Assessment System (ALPAS), for the Denver Public Schools (Colorado). Implementation of the ALPAS system has produced benefits in three major areas: (1) feedback, including progress reports for parents, teachers, and building level administrators; (2) curriculum continuity, due to constraints imposed by the tests on extreme variation in course content; and (3) student achievement. The experiences of the Denver Public Schools in developing ALPAS clearly indicate that an individual district can generate an assessment system of its own that is at least equal to standardized tests in psychometric properties, and possibly superior in its coverage of content taught in the classrooms. In addition, the district-made system has spin-off benefits in the areas of instruction, planning, and student and parent involvement. References are included. (TE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED256039

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

[X] This document has been reproduced as received from the person or organization originating it.
[ ] Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

John G. Conyers

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

DEVELOPING DISTRICT MADE CRITERION REFERENCED TESTS:

A STANDARD OF EXCELLENCE FOR EFFECTIVE SCHOOLS

by

Dr. John G. Conyers
Dr. Kenneth Andrews
Denver Public Schools

Dr. Robert J. Marzano
Mid-continent Regional Educational Laboratory

EA 017 664

1984

Criterion-Referenced Testing Abstract

In this article, we report on the efforts of one district to develop a criterion-referenced testing system and their success, from a number of perspectives, with that system. Specifically, we report on the development of the Academic Learning Progress Assessment System (ALPAS) for the Denver Public Schools.

Much research has been done in recent years on the lack of correlation between standardized testing instruments and school district curriculum goals. Conclusions from this research have suggested school district goals might be better met by supplementing standardized tests with district-made criterion-referenced testing systems. Such a system has been developed with statistically measured success by the Denver Public Schools.

Implementation of the ALPAS system has produced benefits in three major areas: 1) feedback, 2) curriculum continuity, and 3) student achievement.

The feedback capabilities of ALPAS produce substantial benefits, such as, progress reports for parents, teachers, and building level administrators. These reports can be broken down to the pupil, classroom, or school levels. This reporting system has made parents feel more informed as to the progress of their children's education, has provided information with which teachers can assess the competence of their students and plan appropriate instruction, and has provided information to students which allows them to identify specific skill strengths and weaknesses.

Relative to curriculum continuity, ALPAS has been especially impactive. It provides a common thread tying all buildings together. This mitigates against extreme variation in content due to selection by individual teachers while still allowing teachers the freedom to supplement the curriculum in ways they see fit.

One of the more interesting effects of the ALPAS system is its apparent influence on student achievement as measured by standardized testing. Data for only a small proportion of the schools piloting ALPAS was available, but they showed substantially higher percentile gains than the non-pilot schools.

The experiences of the Denver Public Schools in developing the Academic Learning Progress Assessment System clearly indicate that an individual district can generate their own assessment system which is at least equal to standardized tests in psychometric properties and, perhaps, superior in its coverage of content taught in the classrooms. In addition, the district-made system has some spin-off benefits in the areas of instruction, planning, and student and parent involvement. We strongly encourage other districts to consider the possibility of developing their own assessment systems.

INTRODUCTION

Recent research indicates that the extent to which a district's curriculum overlaps with its testing procedures is one characteristic of an effective school (Purkey and Smith, 1982; Glenn, 1981; Veneszy and Winfield, 1979; Mann, 1983). Curriculum/test alignment is considered so important that it is an integral part of the definition of academic learning time (ALT). For example, Berliner (1984) defines ALT as time engaged with materials or activities related to the outcome measure being used (most commonly an achievement test) during which a student experiences a high success rate (p. 60). Implicit in the assertion of the importance of curriculum/test alignment is the assumption that one of the goals of education is to improve pupil performance on standardized tests. About this issue, the late Ron Edmonds wrote: "One of the most important shared characteristics of school improvement programs is their attempt to improve pupil performance on standardized measures of achievement. There are, of course, other important outcomes of schooling that are not measured by standardized achievement tests; however, improved academic achievement undergirds and advances pupil prospects for gain in the more exalted purposes of education." (Edmonds, 1982, p. 4)

Given the importance of testing as a yardstick for determining effectiveness of individual students, schools and districts the issue of the validity of standardized tests as representative indicators of academic achievement becomes a fundamental question. Here we do

not mean "technical validity" as would be measured by a coefficient of congruence between a test and a criterion. Rather, we mean practical validity as discussed by Lazarus (1977). As stated by Lazarus, a test can be objective technically valid, properly standardized and still be a very bad test. Recent research and theory indicate that standardized tests may not be representative indicators of academic ability. Anderson (1972) maintains that test theory overlooks the basic requirements of a system of measurement, "namely that there is a clear and concise definition of the things being counted" (p. 45). Wardrop in his study of standardized tests (1976) noted that items differ so markedly that the operational definition used for the construct being measured seems to be a function of the test author's idiosyncratic feelings about the construct. Only in a few isolated uses is there evidence that the test author made an attempt to underpin item development with construct theory.

All of this implies that the educational system is being driven by reliance on an imprecise measurement system. Some of the classic arguments for this implication have been reported by Houts (1977). There is also an implication that an individual district is left with one course of action - the development of a district-made criterion - referenced testing system. Roid and Haladyna (1980) write that "dissatisfaction with the imprecision produced by item writers who create different items from the same objective has prompted some to reject objective-based tests in favor of other purer forms of criterion-

referenced tests" (p. 310). Popham (1978) states, "The thrust of the emerging criterion-referenced measurement technology. . . is on increasing the capabilities of criterion-referenced tests to produce lucid descriptions of examinees' performance" (p. 91).

A common reaction from a district contemplating the development of their own criterion-referenced testing system is that the final product might be inferior in psychometric properties to standardized tests. In this article we report on the efforts of one district to develop a criterion-referenced testing system and their success, from a number of perspectives, with that system. Specifically, we report on the development of the Academic Learning Progress Assessment System (ALPAS) for the Denver Public Schools.[1] This article is meant as: 1) a description of the ALPAS development process; 2) a report on the quality of the ALPAS system vs. standardized tests; and 3) a report on the effects of the system on various sectors of the district.

THE ALPAS DEVELOPMENT PROCESS

The initial step in the development process was to identify the valued learning objectives of the district. Previous studies of the content of the district's standardized test (The Comprehensive Test of Basic Skills) had indicated that there was a low level of congruence between CTBS and the stated curriculum of the district. Previous studies had also indicated that the curriculum implemented in the classroom was possibly very different from that stated in the curriculum guides.

---

[1]The ALPAS system was developed under the direction of Mr. Ken Andrews.

This is because the final arbiter of what it is that gets taught in the classroom is the classroom teacher. Even though specific curriculum content may be mandated by the district or even the state, teachers will select content based on personal discretion (Fisher, et al., 1980; Schwille, et al., 1981; Carew and Lightfoot, 1979). In light of this, an extensive survey of teachers in the district was conducted to identify the objectives that were being taught. This produced a comprehensive list of content objectives and accompanying scope and sequence. Items were then constructed for each objective at each grade level.

In 1980-1981, ALPAS was pilot tested in 17 elementary schools; in 1981-1982, it was piloted in 38 elementary schools. Items were selected for the final version based on their psychometric properties.

In its final format, ALPAS has a number of unique characteristics:

1. Placement tests in each of the content areas for each grade level (K-6);
2. Pre and posttests for each grade level for each content area;
3. Six interval tests for each content area at each grade level;
4. Objective guides and skill illustrations for each interval;
5. A computer scoring system which provides a variety of descriptive reports for pupils, teachers, and parents;
6. Homework work sheet masters for each subskill area.

COMPARISON WITH STANDARDIZED TESTS

In general, the quality of a standardized test is judged by the standardization or norming group, the reliability as measured by various indices and estimates of various types of validity.

One of the major sources of error in a standardized test is the lack of representativeness of the sample on which the test was normed. To be appropriate, a norm group must be recent and representative (Mehrens and Lehmann, 1980). Recency refers to how long a period of time has elapsed since the norms were developed. Given that the standardized testing industry is a multi-million dollar business, test makers are reluctant to engage in any changes in their product (tests) which will decrease profits (Kohn, 1977). Renorming a test is one of the major expenses in test development. Mehrens & Lehmann (1980, p. 87) warn: "the recency of the norm groups on a published test cannot be judged by the copyright date of the test manual. Any change in the test manual allows the published to revise the copyright date of the manual although the later date may not be an indication of the recency of the norms."

Representativeness refers to the extent to which the norming group includes independent variables (e.g., age, sex, socioeconomic status, race, size of community and geographic location of the subject) which are characteristic of the group being tested. A large sample size is not sufficient. If a sample is biased, making it large does not solve the problem. A common form of sample bias is to sample by convenience rather than by a particular technical sampling procedure (e.g., stratified, random sampling). For example, schools which were cooperative in the norming of one form of the test are commonly contacted for renorming. Unfortunately, many test makers do not report the sampling procedure used, only the number of subjects who participated in the norming.

Because ALPAS is a criterion-referenced system, it cannot be compared with standardized norm referenced tests in a straightforward fashion; however, some comparison can be made on the dimension of norming sample. Millman (1974) discusses the use of statistical procedures to develop the criteria for success or failure on the subtests of a criterion-referenced test. If statistical procedures are used, as they were with ALPAS, then the sample which was used to set criterion scores becomes analogous to a norming sample and faces the same sources of error. In this sense, we might compare the sample used to pilot ALPAS with the sample used to norm a standardized test. This pilot or "reference group" for ALPAS is both recent and representative. Because test results are scored by computer, a cumulative record of student performance can be produced. In effect, criterion scores can be constantly monitored using the results from successive administrations. ALPAS is representative because it obtains data from the entire population of students, an accomplishment impossible for a standardized test.

The ALPAS pretest, posttests, placement tests and interval tests were subjected to standard item analysis procedures. Items that did not discriminate well between high and low achieving students were deleted along with items with exceptionally low or high difficulties. Cranbach's coefficient alpha was used to determine eternal consistency for all tests. These coefficients ranged from .67 to .97 with an average

alpha of .91. This compares well with standardized tests and, in fact, is slightly higher than coefficients commonly found in some standardized tests.

For ALPAS, both content and construct validity were studied. Content validity is related to how adequately the content of and responses to the test samples the domain about which inferences are to be made (Mehrens & Lehmann, 1980, p. 63). Lennon (1956, p. 294) defines content validity as the "extent to which a subject's responses to the items of a test may be considered to be a representative sample of his responses to a real or hypothetical universe of situations which together constitute the area of concern to the person interpreting the test." For standardized tests, item types are commonly identified via a "thorough study and analysis of syllabi, textbooks and programs throughout the country" (Mehrens & Lehman, 1980, p. 159). As previously alluded to, this "paper" curriculum is commonly not the curriculum taught in the classroom. For example, Fisher, et al. (1980) after observing one elementary school teacher for over 90 days found that she taught nothing about fractions despite the fact that the topic was found in the textbook, the district curriculum guide and was mandated by the state. In light of this, it seems logical that for a test to truly sample from the general domain of the content being taught individual teachers must be consulted. As stated previously, identification of the ALPAS objectives was accomplished in part by a thorough survey of content area teachers.

Construct validity is the degree to which test scores can be accounted for by certain explanatory constructs in psychological theory.

Commonly construct validity is studied via factor analysis - the theory being that if the constructs purportedly measured by the test do in fact exist, items in the test which measure similar constructs will tend to group together. Factor analysis is a method for determining the number and nature of the underlying variables among larger numbers of measures. As Kerlinger (1973, p. 659) states, factor analysis "reduces the multiplicity of tests and measures to greater simplicity." The ALPAS tests were subjected to a variety of factor analytic techniques. specifically, initial factor solutions with oblique and orthogonal rotations were used. The results indicated not a clear pattern of factor loadings but a general clustering of items in the expected manner.

THE EFFECTS OF ALPAS

Implementation of the ALPAS system appears to have produced benefits in three major areas: 1) feedback, 2) curriculum continuity, and 3) student achievement. We will briefly consider each.

The feedback capabilities of ALPAS appear to produce substantial benefits. ALPAS generates progress reports for parents, teachers and building level administrators. Reports can be broken down to the pupil, classroom, or school levels. Figure 1 illustrates a pupil progress report for parents. Figure 2 illustrates a classroom report for teachers regarding specific mathematics objectives.

Figure 1

Pupil Progress Report for Parents

Date Tested	09/29	11/04	01/13	01/05
Level/Interval Tested	4-0	4-1	4-2	4-3
School:	283	283	283	283
SKILL DESCRIPTION-MATH				
Extend Number Sequence 99-999	R	R	+	+
Know Place Value Through 10,000's	+	+	R	R
Know Extended Notations 10,000's	+	R	R	+
Master Roman Numerals to 50	-	R	+	+
Add: To 3 Addends, 4 Digits - Regroup	+		+	+
Subtract: To 4 Digits - Regrouping	+		+	R
Master 100 Mult., Division Facts	R			+
Multiply: 2 Digits x 1-2 Digits	-			
Divide: 2-3 Digits by 1 (Even)	-			+
Use Zero with All 4 Operations	+			
Match No. Line Point to Mixed No.	-			
Money: Make Change Up to $10.00	+	+	R	
Money: Add, Subtract, Multiply-$10.00	-	R	+	
Measure or Convert Dry Weight	-		-	-
Measure, Convert Liquid Measures	R		R	R
Measure, Linear Distrance	+		R	-
Find Perimeter: Metric, Customary	-			R
Round Off Numbers to Nearest 100	R			
Calculate Averages	-			-
Solve Word Problem: 1-2 Step, 4 OP	R			
Use Decimals with Money: To $100	R	R	+	
Identify Right Angle, Vertex, Rays	R			R
Know Temperature on a Thermometer	-			
**Total Mastered	7	2	6	5

Key: + = Mastery of Skill (80-100% Correct)
R = Review needed (50-70% Correct)
- = Not Mastered (1-49% Correct)
Blank = Not Tested at this Interval

Figure 2

Class Graph by Test Item

ITEM DESCRIPTOR	NO. OF PUPILS PASSING THE ITEMOUT OF 23 TESTED	
1. Multiply: To 3 Place Multiplier	XXXXXXXXXXXXXXXXXX	(20)
2. Multiply Decimals: Annex Zeros	XXXXXXXXXXXXXXXXXX	(19)
3. Subtract: To 5 Digits, Renaming	XXXXXXXXXXXXXXXXXXXX	(22)
4. Subtract: To 5 Digits, Renaming	XXXXXXXXXXXXXXXXXXXX	(22)
5. Divide Decimal: Dividend Only	XXXXXXXXXXXXXXXXXX	(19)
6. Divide: To 5 Numbers by 2 Numbers	XXXXXXXXXXXXXXXXXX	(19)
7. Divide: To 5 Numbers by 2 Numbers	XXXXXXXXXXXXXXXXXX	(19)
8. Add: To 5 Digits, 3-4 Addends, Rename	XXXXXXXXXXXXXXXXXXXX	(22)
9. Add, Subtract Decimals - Hundredths	XXXXXXXXXXXXXXXXXXXX	(22)
10. Add: To 5 Digits, 3-4 Addends, Rename	XXXXXXXXXXXXXXXXXX	(19)
11. Convert: Improper Fract./Mixed No.	XXXXXXXXXXXXXXXXXXX	(21)
12. Convert: Improper Fract./Mixed No.	XXXXXXXXXXXXXXXXXXXXX	(23)
13. Convert Fraction/Decimal/Ratio	XXXXXXXXXXXXXXXXXXX	(21)
14. Convert Fraction/Decimal/Ratio	XXXXXXXXXXXXXXXXXX	(20)
15. Add, Subtract Decimals - Hundredths	XXXXXXXXXXXXXXXXXXXX	(22)

This reporting system has: 1) made parents feel more informed as to the progress of their children's education; 2) provided information with which teachers can assess the competence of their students and plan appropriate instruction; and 3) provided information to students which allows them to identify specific skill strengths and weaknesses. Parental involvement has been shown to be a contributing factor to school effectiveness (Armor, et al., 1976; Coleman, et al., 1981; Levine & Stark, 1981) although findings are mixed on this issue (Purkey & Smith, 1982). Proper assessment of student abilities has been cited as a key component of effective instruction by a number of researchers and theorists (Good & Grouws, 1979; Gage & Giacini, 1981; Rosenshine, 1983; Hunter, 1984; Berliner, 1984). Similarly, systematic feedback to students about their progress in specific skill areas has been shown to positively affect achievement (Cohen, 1981). In short, the feedback capabilities of ALPAS impact the district not just in record keeping but in planning, instruction, student and parental involvement.

Relative to curriculum continuity, ALPAS has been especially impactive. As is the case with most large districts, the Denver Public Schools allows individual buildings to select basal materials from an approved set. Although this allows for the expression of individual differences in philosophy and teaching pedagogy at the building level, it also creates the potential for discontinuity in terms of districtwide curriculum. ALPAS provides a common thread tying all buildings together. This mitigates against extreme variation in content due to selection

by individual teachers. It is important to note here that ALPAS does not stop selection of content by individual teachers; it simply ties all teachers to a common core of content. Teachers are still free to supplement the curriculum in ways they see fit.

One of the more interesting effects of the ALPAS system is its apparent influence on student achievement as measured by a standardized test (CTBS). Data for only a small proportion of the schools piloting ALPAS was available for only two grade levels and for only one subject area (math). Hence, its representativeness is suspect. However, if we compare the gains (in percentile rank) of the pilot schools with the non-pilot schools for the first and second years of the ALPAS implementation some interesting trends emerge. Table 1 displays these data.

TABLE 1

Pilot vs. Non-Pilot Math Percentile Gain Scores

First Year of ALPAS Implementation

	Grade 2	Grade 5
Pilot school gain	4.61 SD = 5.37 N = 5	4.51 SD = 11.89 N = 12
Non-pilot school gain	.89 SD = 10.08 N = 45	2.56 SD = 8.20 N = 36

Second Year of ALPAS Implementation

Pilot school gain	5.21 SD = 4.87 N = 5	11.42* SD = 12.31 N = 12
Non-pilot school gain	3.24 SD = 9.82 N = 45	-2.86 SD = 9.52 N = 36

(Note: * indicates differences are significant at the .05 level when percentiles are transformed to standard scores and a comparison made using independent t-tests)

Table 1 indicates that the ALPAS pilot schools had higher percentile gains in all four comparisons (pilot vs. non-pilot schools for grades two and five for each of the two implementation years). In the second year of implementation, the non-pilot schools actually went down 2.86 percentile points in fifth grade math achievement while the ALPAS pilot schools showed a gain of 11.42 percentile points. Again given the scarcity of data, we cannot generalize from these results but the fact that in all four cases the ALPAS schools exhibited higher gains does raise an intriguing possibility - namely, that ALPAS had an effect on student achievement as measured by a standardized test. If this were the case, an explanation might be that ALPAS' effect on the areas previously mentioned (e.g., teacher assessment and planning, student feedback, parental involvement, curriculum continuity) improved teaching and learning to the extent that the improvement could be detected by measures not referenced to the ALPAS system (e.g., norm referenced standardized tests).

CONCLUSIONS

The experiences of the Denver Public Schools in developing the Academic Learning Progress Assessment System clearly indicate that an individual district can generate their own assessment system which is at least equal to standardized tests in psychometric properties and perhaps superior in its coverage of content taught in the classrooms. In addition, a district-made system has some spin-off benefits in the

areas of instruction, planning, and student and parent involvement. We should note here that we do not believe that district-made tests should necessarily replace norm referenced standardized tests. Rather, they should be used as a supplement and complement to standardized tests. We strongly encourage other districts to consider the possibility of developing their own assessment systems.

REFERENCES

Anderson, R.C. How to Construct Achievement Tests to Assess Comprehension, Review of Educational Research, 1972, 42, 145-170.

Armor, D., Conry-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., Pauly, E., & Zellerman, G. Analysis of the School Preferred Reading Program in Selected Los Angeles Minority Schools, Santa Monica, California, Rand Corporation (EDRS # ED 130 243), 1976.

Berliner, D. A Glass Half-full: A Review of Research in Teaching. In P. Hosford (ed.) Using What We Know About Teaching, Alexandria, Virginia, Association for Supervision and Curriculum Development, 1984.

Carew, J. & Lightfoot, S.L. Beyond Bias, Cambridge, Massachusetts, Harvard University Press, 1979.

Cohen, M. Effective Schools Research: Towards Useful Interpretations. Paper presented at the American Association of School Administrators Instructional Leadership Conference, Summer, 1981.

Coleman, J.S., Hoffer, T., & Kilgore, S. Public and Private Schools. Chicago, University of Chicago, National Opinion Research Center, 1981 (EDRS # ED 197 503).

Edmonds, R. Programs of School Improvement: An Overview. Paper presented at Conference on Research on Teaching, Airlee House, Virginia, 1982.

Fisher, C.W., Berliner, D.C., Filby, N.N., Marliave, R.S., Cahen, L.S., & Dishaw, M.M. Teaching Behaviors, Academic Learning Time and Student Achievement: An Overview. In C. Denham & A. Lieberman (eds.) Time to Learn, Washington, D.C., U.S. Department of Education, National Institute of Education, 1980.

Gage, N.L. & Giaconia, R. Teaching Practices and Student Achievement: Casual Connections, New York University Education Quarterly, 113, 1981, 2-9.

Glenn, D.C. What Works? An Examination of Effective Schools for Poor Black Children, Cambridge, Massachusetts, Center for Law and Education, Harvard University, 1981.

Good, T.L. & Grouws, D. The Missouri Mathematics Effectiveness Project: An Experimental Study in Fourth Grade Classrooms; Journal of Educational Psychology, 1979, 71, 355-362.

Houts, P.L. (ed.) The Myth of Measurability, New York, Hart Publishing Company, 1977.

Hunter, M. Knowing, Teaching and Supervising. In P. Hosford (ed.) Using What We Know About Teaching, Alexandria, Virginia, Association for Supervision and Curriculum Development, 1984.

Kerlinger, F. Foundation of Behavioral Research, New York, Holt, Rinehart & Winston, 1973.

Kohn, S.D. The Numbers Game: How the Testing Industry Operates. In P.L. Houts (ed.) The Myth of Measurability, New York, Hart Publishing Company, 1977.

Lazarus, M. Coming to Terms with Testing. In P.L. Houts (ed.) The Myth of Measurability, New York, Hart Publishing Company, 1977.

Lennon, R.T. Assumptions Underlying the Use of Content Validity, Educational and Psychological Measurement, 1956, 16, 294-304.

Levine, D.U. & Stark, J. Instructional and Organizational Arrangements and Processes for Improving Academic Achievement at Inner City Elementary Schools: A Study of Chicago Mastery Learning Reading Programs and Other Schoolwide Approaches for Improving Reading at Selected Schools in Chicago, Los Angeles and New York: Extended Summary and Conclusions. Paper presented at the annual meeting of the American Educational Research Association, New York, 1982 (EDRS # ED 213 814).

Mann, D. A Delphi Analysis of the Instructionally Effective School, New York, Teacher College, Columbia University, 1983.

Mehrens, W.A. & Lehmann, I.J. Standardized Tests in Education, New York, Holt, Rinehart & Winston, 1980.

Millman, J. Criterion-referenced Measurement. In W.J. Popham (ed.) Evaluation in Education: Current Applications, Berkeley, California, McCutchan Publishing Company, 1974.

Popham, W.J. Criterion-referenced Measurement, Englewood Cliffs, New Jersey, Prentice-Hall, 1978.

Purkey, S. & Smith, M. Effective Schools - A Review. Paper presented at the Conference on Research in Teaching, Airlee House, Virginia, 1982.

Roid, G. & Haladyna, J. The Emergence of an Item-writing Technology, Review of Educational Research, 1980, 50, 293-314.

Rosenshine, B. Teaching Functions in Instructional Programs. The Elementary School Journal, Vol. 83, 1983, 335-352.

Schwille, J., Porter, A., Belli, A., Floden, R., Freeman, D., Kappen, L., Kuhs, T., & Schmidt, W.J. Teachers as Policy Brokers in the Content of Elementary School Mathematics, East Lansing, Institute for Research on Teaching, Michigan State University, 1981.

Venezky, R.L. & Winfield, L.F. Schools that Succeed Beyond Expectations in Teaching Reading, Newark, Delaware, College of Education, University of Delaware, 1979.

Wardrop, J.L. Standardized Lists in the Schools: Uses and Roles, Monterey, California, Brooks-Cole, 1976.